RADemics

Multi-Stage Threat Analysis with Hybrid Machine Learning Models Combining Static and Dynamic Data Features

Babeetta Bbhagat, J. Rohini

MIT ADT UNIVERSITY, SENGUNTHAR ENGINEERING COLLEGE

Multi-Stage Threat Analysis with Hybrid Machine Learning Models Combining Static and Dynamic Data Features

[1]Babeetta Bbhagat, Assistant Professor, Department of Computer Science and Engineering, MIT ADT University, Loni Kalbhor, Maharashtra, India. Babitas12@gmail.com

[2]J. Rohini, Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Erode, India.rohinime1997@gmail.com

Abstract

This chapter explores the advancements in multi-stage threat detection systems utilizing hybrid machine learning models that integrate both static and dynamic data features. It delves into the significance of combining these data types for superior threat classification, focusing on the challenges and solutions in handling high-dimensional feature spaces and real-time data processing. By leveraging advanced feature selection, hybrid classification algorithms, and real-time behavior analysis, the chapter provides insights into improving detection accuracy and system scalability. Emphasizing performance evaluation methods such as cross-validation and real-world testing, it highlights the importance of assessing model effectiveness across multiple stages in dynamic cybersecurity environments. The integration of static and dynamic data presents a powerful framework for detecting and mitigating emerging threats, offering valuable perspectives for future research and application in real-time threat analysis.

Keywords: Multi-Stage Threat Detection, Hybrid Machine Learning, Static Data, Dynamic Data, Feature Selection, Real-Time Analysis.

Introduction

In the rapidly evolving field of cybersecurity, multi-stage threat detection systems are becoming increasingly essential [1]. Traditional threat detection methods often fall short in addressing the complexity and variety of modern threats [2]. The integration of hybrid machine learning models combining static and dynamic data features provides a robust approach to overcoming these limitations [3]. Static data features, such as network configurations, historical logs, and device attributes, offer a steady foundation for threat detection, while dynamic data, such as real-time user behavior, system interactions, and network traffic, captures the evolving nature of attacks [4-6]. This hybrid approach not only improves the accuracy of detection but also enhances the ability to adapt to novel and sophisticated threats thatnot be detected by traditional systems [7,8]. The fusion of these data types offers a comprehensive view of the system's state, enabling more accurate and timely responses to emerging threats [9,10].

One of the primary challenges in multi-stage threat detection systems lies in the integration of static and dynamic features [11]. Static features are often limited in their ability to respond to the continuously changing landscape of cybersecurity threats [12]. On the other hand, dynamic features provide real-time insights but can lead to high-dimensional data thatoverwhelm traditional machine learning models [13-16]. Therefore, effective feature selection and extraction techniques are critical to reducing the complexity and ensuring that only the most relevant information was processed [17]. These systems must be designed to scale efficiently, managing both the volume and the variety of data while maintaining real-time responsiveness [18]. Addressing these challenges requires a sophisticated combination of feature engineering, model optimization, and computational power to ensure that multi-stage systems can handle both types of data without sacrificing performance [19,20].

Hybrid machine learning models are at the core of modern multi-stage threat detection systems [21]. These models combine the strengths of various algorithms to address the complexities of both static and dynamic data features [22]. For example, decision trees and random forests can handle static data effectively, while more advanced models like recurrent neural networks (RNNs) or convolutional neural networks (CNNs) are adept at processing dynamic, time-series data [23-25]. By leveraging ensemble methods, these hybrid models can improve detection accuracy by capturing diverse data patterns and detecting subtle anomalies thatotherwise go unnoticed.